

04014

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-338646

(43)Date of publication of application : 10.12.1999

(51)Int.Cl.

G06F 3/06

G06F 3/06

(21)Application number : 10-147552

(71)Applicant : HITACHI LTD

(22)Date of filing : 28.05.1998

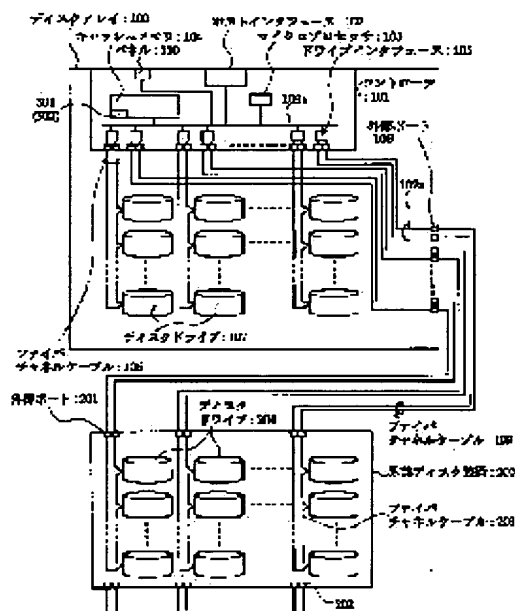
(72)Inventor : HIREZAKI KATSUMI

(54) DISK ARRAY DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To realize remote mirroring of a disk array device at low cost.

SOLUTION: Disk drives 107 inside a disk array device 100 are connected to drive interfaces 105 of a controller 101 through fiber channel cables 106 to construct an FC-AL, also, disk drives 204 in an external disk drive 200 at a remote place are connected to drive interfaces 105 of the controller 101 through external ports 108 and fiber channel cables 109 to construct an FC-AL. Disaster recovery mirroring which stores the same data in a multiple way between a disk array of the local disk drives 107 and a disk array of the remote disk drives 204 is realized in the common controller 101.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-338646

(43)公開日 平成11年(1999)12月10日

(51)Int.Cl.⁸

G 0 6 F 3/06

識別記号

3 0 4

5 4 0

F I

G 0 6 F 3/06

3 0 4 B

5 4 0

審査請求 未請求 請求項の数 3 O L (全 8 頁)

(21)出願番号

特願平10-147552

(22)出願日

平成10年(1998)5月28日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 鏑崎 克巳

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74)代理人 弁理士 筒井 大和

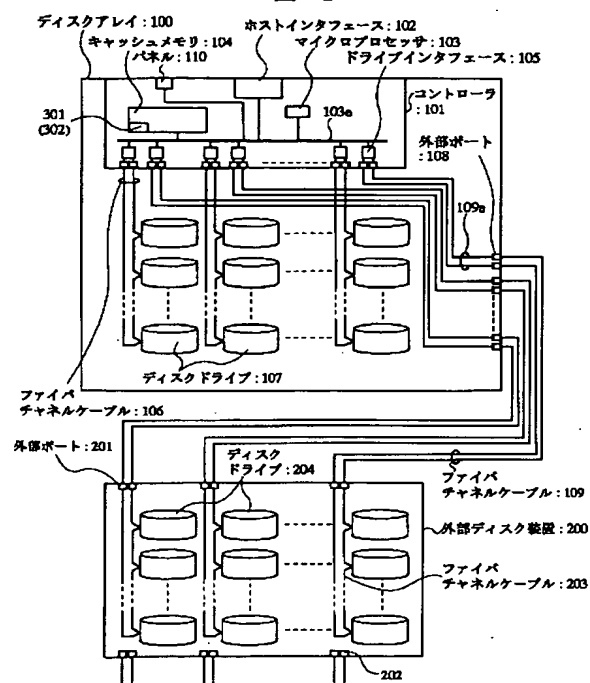
(54)【発明の名称】 ディスクアレイ装置

(57)【要約】

【課題】 ディスクアレイ装置の遠隔ミラーリングを安価に実現する。

【解決手段】 ディスクアレイ装置100の内部のディスクドライブ107を、ファイバチャネルケーブル106を介してコントローラ101のドライバインタフェース105に接続してFC_ALを構成するとともに、外部ポート108、ファイバチャネルケーブル109を介して、遠隔地の外部ディスク装置200内のディスクドライブ204を、コントローラ101のドライバインタフェース105に接続してFC_ALを構成し、共通のコントローラ101にて、ローカルのディスクドライブ107のディスクアレイと、リモートのディスクドライブ204のディスクアレイとの間で同一データを多重に格納するディザスタリカバリ遠隔ミラーリングを実現する。

図 1



【特許請求の範囲】

【請求項 1】 複数の単体ドライブと、前記単体ドライブの動作を制御するコントローラとを含むディスクアレイ装置であって、

前記単体ドライブは、前記ディスクアレイ装置の内部に配置され、前記コントローラと複数の第 1 のインタフェースにて接続される第 1 の単体ドライブと、前記ディスクアレイ装置の外部に配置され、前記ディスクアレイ装置に設けられた外部接続ポートを介して、前記コントローラと複数の第 2 のインタフェースにて接続される第 2 の単体ドライブと、

を含むことを特徴とするディスクアレイ装置。

【請求項 2】 複数の単体ドライブと、前記単体ドライブの動作を制御するコントローラとを含むディスクアレイ装置であって、

前記単体ドライブは、前記ディスクアレイ装置の内部に配置され、前記コントローラと複数のインタフェースにて接続される第 1 の単体ドライブと、

前記ディスクアレイ装置の外部に配置され、前記ディスクアレイ装置に設けられた外部接続ポートを介して、前記インタフェースに接続される第 2 の単体ドライブと、を含むことを特徴とするディスクアレイ装置。

【請求項 3】 請求項 1 または 2 記載のディスクアレイ装置において、

前記インタフェースは、ファイバチャネル・アービトラレーテッド・ループ (FC_{AL}) であり、前記コントローラは、同一のデータを、前記第 1 および第 2 の単体ドライブに多重に格納することを特徴とするディスクアレイ装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、ディスクアレイ技術に関し、特に、ディスクアレイにおけるデータ多重化技術等に適用して有効な技術に関する。

【0002】

【従来の技術】 コンピュータシステムの規模は、年々大規模化してきており、それに伴い扱うデータ量も増大してきている。このデータ量増大に伴い、大容量記憶システムとして耐故障性を考慮し、ディスクアレイ装置を使用する傾向に有る。これはディスクドライブが突然起こす故障に対して、データ保証を行う技術である。このディスクアレイ技術の中の一つに、ミラーリングという方法が有る。ミラーリングとは、複数台のディスクドライブのそれぞれに対し、同じデータを二重に書き込む手法である。このミラーリングの手段には次のようなものが存在する。

【0003】 (1) ディスクアレイ装置を RAID 1 で使用する。

【0004】 (2) ソフトウェア RAID を使用し、別々のディスクに同じデータを書き込む。

【0005】 (3) たとえば、特開平 8-272666 号公報に開示された技術のように、ネットワークを使用し、ローカルのディスクアレイ装置に書き込んだデータをリモートのディスクアレイ装置にも書き込む。

【0006】 前記 (1)、(2) は、ローカルにてデータの二重化を行うものである。これは、単体のディスクドライブの故障によるデータ喪失防止を行うことは可能であるが、地震や火事等といった災害が発生した場合、二重化したディスクドライブが、共に破壊される可能性があり、データを二重化していてもデータを喪失する可能性が非常に高い。

【0007】 一方、(3) はネットワークを使用しているため、遠隔地に二重化したデータが存在するので、災害によりローカルのディスクドライブが全て破壊されたとしても、遠隔地のディスクからデータ復旧が可能であるが、遠隔地に対してもディスクアレイ装置を必要とするため、コストがかかる。

【0008】

【発明が解決しようとする課題】 上述のように、ディスクアレイ装置のミラーリングは、アレイ内のディスクにて行うため、災害発生時にミラーリングしたデータも同時に破壊され、データ喪失に至る、という技術的課題がある。

【0009】 また、遠隔ミラーリングをディスクアレイ装置にて行う場合、中央処理装置、もしくは、ディスクアレイ装置がネットワークを使用して、遠方のディスク装置にデータを転送するが、その際は遠方のディスク装置に対しても、別のディスクアレイ装置が必要となり、ディスクアレイにおけるデータ二重化のコストが高くなる、という技術的課題もある。

【0010】 本発明の目的は、災害等によるデータ喪失を確実に防止することが可能なディスクアレイ技術を提供することにある。

【0011】 本発明の他の目的は、安価に、遠隔ミラーリングを実現することが可能なディスクアレイ技術を提供することにある。

【0012】

【課題を解決するための手段】 本発明では、ディスクアレイ装置において、装置内部に配置されたローカルの第 1 の単体ドライブ、およびディスクアレイ装置の外部に設けられたリモートの第 2 の単体ドライブを、アレイコントローラに対してファイバチャネル (FC_{AL}: Fibre Channel-Arbitrated Loop) 等のインタフェースにて接続するものである。

【0013】 より具体的には、一例として、ディスクアレイ装置にファイバチャネル用制御ポートを設け、遠隔地にあるディスク装置 (ドライブ群) にも、ファイバチャネル用制御ポートを設けて、お互いをファイバチャネルプロトコルを使用して接続する。

【0014】 ファイバチャネル用制御ポート数、および

そのポートにFC__AL接続するドライブの数は、ローカルのディスクアレイのポート数、およびポートに付属のドライブ数と同数とし、ローカルのディスクアレイの構成と同様にする。

【0015】ファイバチャネルプロトコルにより、遠隔地のドライブに、ローカルのディスクアレイ装置を接続可能であり、かつアレイコントローラ等の特別な制御装置を遠隔地のディスク装置に持つ必要がなく、安価にミラーリングを行うことが可能である。

【0016】ファイバチャネルプロトコルによりお互いを接続することで、ネットワークを使用せず遠方にあるミラーリング用ドライブに接続可能であり、かつ特別な制御装置をミラーリング用ドライブ群側に持つ必要がなくなり、安価に遠隔ミラーリングを行うことが可能となる。

【0017】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

【0018】ミラーリングとは、データの二重化のことである。本実施の形態のディスクアレイ装置では、主ドライブと副ドライブを定義する。上位装置から発行されたデータは、主ドライブに書き込むと同時に副ドライブにも書き込まれる。ドライブからのデータ読み出しは、基本的に主ドライブから行うが、主ドライブが何らかの故障によりアクセス不能となった場合、副ドライブからデータを読み出す。このような方法で、ディスクドライブの障害によるデータロス防止を行うことをミラーリングといい、特に、ディスクアレイではRAID1と呼ばれる。このRAID1は、ディスクアレイ内部のディスク群を主および副として使用するため、火事や地震といった災害などによりディスクアレイ装置そのものが壊れた場合データを失う。

【0019】そこで、本発明のディスクアレイ装置の一実施の形態であるディザスタリカバリ遠隔ミラーリングシステムは、この技術的課題を解決するために、以下のように、ファイバチャネルプロトコルを使用し、遠隔地に設置されたディスクドライブに対してミラーリングを行うものである。

【0020】図1は、本発明のディスクアレイ装置の一実施の形態であるディザスタリカバリ遠隔ミラーリングシステムの構成の一例を示す概念図である。

【0021】本実施の形態のディザスタリカバリ遠隔ミラーリングシステムは、ディスクアレイ装置100と外部ディスク装置200から構成される。

【0022】ディスクアレイ装置100は、コントローラ101、外部ポート108、ディスクドライブ107、ファイバチャネルケーブル106から構成される。

【0023】コントローラ101は、ディザスタリカバリ遠隔ミラーリングシステム全体を制御し、上位装置であるホストとディスクアレイ装置100の間のプロトコ

ル制御を司るホストインタフェース102、装置全体の制御を行うマイクロプロセッサ103、ホストからのデータ、および後述の図2に例示されるような制御情報を保持するキャッシュメモリ104、ディスクドライブ107、204とのプロトコル制御を行うドライブインタフェース105から構成される。マイクロプロセッサ103に対して、それ以外の各部はシステムバス103aを介して接続され、マイクロプロセッサ103の制御の下で動作する。

【0024】ディスクドライブ107は、ファイバチャネルケーブル106を使用し、コントローラ101のドライブインタフェース105と接続する。

【0025】外部ポート108は、外部ディスク装置200とファイバチャネルケーブル109を使用して接続するためのインタフェースであり、ファイバチャネルケーブル109aを介して、コントローラ101のドライブインタフェース105と接続されている。

【0026】外部ディスク装置200は、外部ポート201、外部ポート202、複数のディスクドライブ204、ファイバチャネルケーブル203から構成される。外部ポート201は、ディスクアレイ装置100とディスクドライブ204間のインタフェースであり、ディスクドライブ204はファイバチャネルケーブル203を使用してディスクアレイ装置100と接続される。

【0027】なお、外部ポート202は、増設用に用いられ、たとえばデータを3重以上に多重化する場合に、たとえば外部ディスク装置200と同様の構成を持つ外部ディスク装置が、必要に応じて接続される。

【0028】次に、本実施の形態にて用いられる制御情報の一例について示す。図2は本実施の形態で使用する記憶管理体系（制御情報）の一例を示す概念図である。本実施の形態では、一例として図2に例示されるドライブ管理テーブル301と、データ管理テーブル302とを用いる。ドライブ管理テーブル301は、主ドライブおよび副ドライブの各々について、ドライブの数分のエントリを持つ。

【0029】ミラーリングを行う主ドライブ（ディスクドライブ107）と副ドライブ（ディスクドライブ204）は、それぞれ別々のファイバチャネルケーブル106、およびファイバチャネルケーブル109で、コントローラ101のドライブインタフェース105に接続され、独立なFC__ALループを構成する。

【0030】ミラーリングを制御するに際し、まず主ドライブと副ドライブの対応を取る必要が有る。ファイバチャネルにおけるデバイスは、それぞれ世界中で一意のWWN（World Wide Name）を持つ。ディスクドライブ107、ディスクドライブ204のWWNは、FC__ALループ初期化時に、コントローラ101が認識可能であるが、この時、あるディスクドライブが、ディスクアレイ装置100に属しているのか、外部ディスク装置2

00に属しているのか、認識不可能である。また外部ディスク装置200は外部ポート202を使用した場合には、複数台接続可能なので、どの外部ディスク装置200にあるのかも不明である。そこで、ディスクアレイ装置100のパネル110からディスクアレイ装置100に属しているディスクのWWNを設定し、ドライブ管理テーブル301のWWN301-1-cおよびWWN301-2-cに登録する。

【0031】また、事前にパネル110を使用し、ループID301-1-b、ループID301-2-bをそれぞれ設定することによって、主側に使用するループと副側に使用するループの登録を行う。

【0032】D_ID301-1-a、D_ID301-2-aはPORT_IDでファイバチャネルでデバイスを特定するのに使用される。これはFC_AL初期化時に動的に決定される。

【0033】ドライブ正常フラグ301-1-e、ドライブ正常フラグ301-2-eは、そのドライブが正常か異常かを示す。異常の場合はそのドライブに対し、I/Oを行わない。副ドライブへのポインタ301-1-d、主ドライブへのポインタ301-2-dはミラーリングを行っているディスクの互いの対応ドライブを認識するためのエントリへのポインタである。

【0034】上位装置からの書込データは、たとえば図2に例示されるデータ管理テーブル302を使用して管理する。データへのポインタ302-aはデータの存在するキャッシュのアドレスである。D_ID302-bはデータの書き込む位置にあるドライブのD_IDである。LBA302-cはデータを反映させるディスク内のアドレスである。主書込未反映フラグ302-dは、このデータが主側のディスクドライブ107に対し既に反映されたかどうかを示す。同様に、副書込未反映フラグ302-eは、このデータが副側のディスクドライブ204に対し反映されたか否かを示す。

【0035】以下、本実施の形態の作用の一例について説明する。

【0036】図3のフローチャートにて、本実施の形態の遠隔ミラーリングの処理の一例を遠隔ミラーリングの制御法の一例として示す。

【0037】上位装置から転送されたコマンドはディスクアレイ制御を行うコントローラ101により解析され、WRITE/READ等が判別される(ステップ401)。

【0038】WRITEコマンドの場合、上位装置から受領したコマンドに対するデータが転送され、このデータをキャッシュメモリ104に保持する(ステップ402)。

【0039】このとき、WRITEコマンドのロジカルブロックアドレス(LBA)およびレングス(データ長)からローカルのディスクアレイ装置100上のディ

スクドライブ107のD_IDとLBA、レングスを算出し、データ管理テーブル302を設定する(ステップ403)。

【0040】次に、キャッシュメモリ104上のデータをディスクドライブ107(204)に書き込む。キャッシュメモリ104上のデータをディスクドライブ107(204)に書き込むタイミングは、たとえば、タイマにより時間を監視し、一定の時間がきたらディスクドライブ107(204)に反映させる、もしくはキャッシュメモリ104の使用エリアの容量に閾値を使用し、閾値を超えたらディスクドライブ107(204)に反映させる、もしくは、I/Oの負荷が大きくないときにディスクドライブ107(204)に反映させるといった形で非同期に行い、不要なディスクドライブ107(204)へのアクセスを減らす。

【0041】書込契機になると、まず、データ管理テーブル302のD_ID302-bをタグとしてドライブ管理テーブル301のD_ID301-1-a、副ドライブへのポインタ301-1-dから、副ドライブの有無を確認し(ステップ404)、ない場合には、主側のディスクドライブ107のみにデータ書込を実行し(ステップ405)、キャッシュメモリ104の領域を解放した後(ステップ406)、ホストにWRITE完了を応答する(ステップ407)。

【0042】一方、副ディスクドライブが存在する場合には、ローカルのディスクドライブ107および対応するミラーリングディスクに対し同時に書込処理を行う。

【0043】すなわち、データ管理テーブル302のD_ID302-bをタグとしてドライブ管理テーブル301のD_ID301-1-a、副ドライブへのポインタ301-1-dから、対応する主/副ドライブを特定した後(ステップ408)、データ管理テーブル302の主書込未反映フラグ302-d、副書込未反映フラグ302-eを共にONにし(ステップ409)、その後、主/副ドライブの両方にデータ書込命令を発行する(ステップ410)。

【0044】主/副ドライブの各々では、それぞれWRITE(データ反映)完了を監視し(ステップ411、ステップ413)、ディスクドライブ107、ディスクドライブ204の各々ではデータを反映すると、反映した側のディスクドライブに対するデータ未反映フラグをOFFとする(ステップ412、ステップ414)。

【0045】その際、対応する他方のディスクのデータ未反映フラグを確認し、OFFとなっていればキャッシュ領域として再びデータを書き込むことを可能とし、ONのままであれば、これがOFFになるまでこのキャッシュ領域の使用を不可とし、ミラーリングが未反映となることを防止する(ステップ415)。

【0046】主/副のディスクドライブ107および204の両方にデータの反映が完了すると、キャッシュメ

モリ104の領域を解放した後(ステップ406)、ホストにWRITE完了を応答する(ステップ407)。

【0047】一方、READコマンドの場合、基本的に主側のディスクドライブ107からデータを取得するが、この主側のディスクドライブ107が故障していた場合、データを取得できないため、この時は対応する副側のディスクドライブ204からデータを取得する。

【0048】すなわち、図4のフローチャートに例示されるように、まず、コマンド受領時にディスクアレイ装置上のディスクのD_IDとLBA、レングスを算出した後(ステップ501)、アクセス対象の主側のディスクドライブ107が正常か否かを、ドライブ管理テーブル301の当該ドライブのドライブ正常フラグ301-1-eを参照して判別し(ステップ502)、正常な場合には、主側のディスクドライブ107からデータを読み出して(ステップ503)、ホスト側に転送する(ステップ504)。

【0049】異常の場合には、データ管理テーブル302のD_ID302-bをタグとしてドライブ管理テーブル301のD_ID301-1-a、副ドライブへのポインタ301-1-dから、対応する主/副ドライブを特定したのち(ステップ505)、特定された副側のディスクドライブ204からデータを読み出して(ステップ506)、ホスト側に転送する(ステップ504)。

【0050】このように、本実施の形態のディザスタリカバリ遠隔ミラーリングによれば、ディスクアレイ装置100内のローカルな複数のディスクドライブ107と、遠隔地の外部ディスク装置200に設けられた複数のディスクドライブ204の各々を、それぞれ別々のファイバチャネルケーブル106、およびファイバチャネルケーブル109で、コントローラ101のドライブインタフェース105に接続して、独立なFC_ALループを構成し、対応するディスクドライブ107とディスクドライブ204に同じデータを格納する遠隔ミラーリングを行うので、中央処理装置等のホストからディスクアレイ装置100へのライトデータのコピーを遠隔地の外部ディスク装置200に保持することが可能となり、災害発生時にディスクアレイ装置100が破壊されても遠隔地の外部ディスク装置200にてデータ復旧が可能となる。

【0051】また、ファイバチャネルプロトコルを使用して、ディスクアレイ装置100内のコントローラ101と、遠隔地の外部ディスク装置200と接続しているため、遠隔地側にある外部ディスク装置200はアレイコントローラを備える必要がなく、安価に遠隔ミラーリングを行うことが可能となる。

【0052】なお、図1に例示した構成では、ディスクアレイ装置100内のローカルのディスクドライブ107と、外部ディスク装置200のディスクドライブ204とは、ファイバチャネルケーブル106の系列のFC

_ALと、ファイバチャネルケーブル109a、外部ポート108、ファイバチャネルケーブル109からなる系列のFC_ALとに独立に接続されていたが、これに限らず、たとえば、図5に例示されるように、ディスクドライブ107が接続されるファイバチャネルケーブル106aを、外部ポート108に接続し、この外部ポート108を介して、ファイバチャネルケーブル109と接続することにより、主側のディスクドライブ107と、外部ディスク装置200内の副側のディスクドライブ204とが、同一系列のFC_ALに属する構成としてもよい。

【0053】そして、ローカルのディスクドライブ107と、対応する外部のディスクドライブ204に同じデータを二重に格納する遠隔ミラーリングを行う。

【0054】この図5の変形例の場合には、図2に例示されたドライブ管理テーブル301で、主/副のディスクドライブの各々が帰属するFC_ALを区別するための、ループID301-1-b、ループID301-2-bのデータは不要であり、制御情報が簡素化されるとともに、ディスクアレイ装置100内のコントローラ101におけるドライブインタフェース105の数を減らすことができ、ハードウェア構成が簡素化される、という利点がある。

【0055】以上本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【0056】たとえばインタフェースとしては、ファイバチャネルに限らず、遠距離のデータ授受が可能なインタフェースであれば、他のインタフェースでもよい。

【0057】

【発明の効果】本発明のディスクアレイ装置によれば、災害等によるデータ喪失を確実に防止することができる、という効果が得られる。

【0058】また、安価に、遠隔ミラーリングを実現することができる、という効果が得られる。

【図面の簡単な説明】

【図1】本発明のディスクアレイ装置の一実施の形態であるディザスタリカバリ遠隔ミラーリングシステムの構成の一例を示す概念図である。

【図2】本発明のディスクアレイ装置の一実施の形態であるディザスタリカバリ遠隔ミラーリングシステムで使用する制御情報の一例を示す概念図である。

【図3】本発明のディスクアレイ装置の一実施の形態であるディザスタリカバリ遠隔ミラーリングシステムの作用の一例を示すフローチャートである。

【図4】本発明のディスクアレイ装置の一実施の形態であるディザスタリカバリ遠隔ミラーリングシステムの作用の一例を示すフローチャートである。

【図5】本発明のディスクアレイ装置の一実施の形態で

あるディザスタリカバリ遠隔ミラーリングシステムの変形例を示す概念図である。

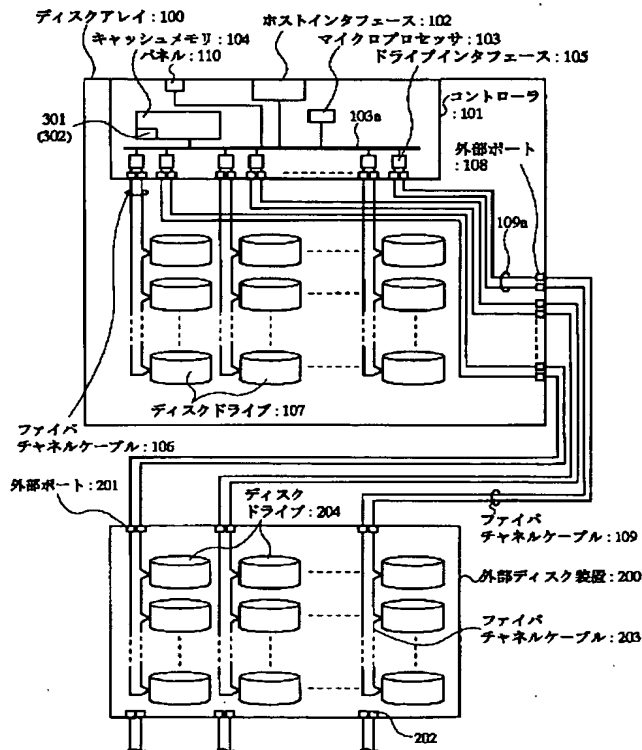
【符号の説明】

100…ディスクアレイ装置、101…コントローラ、102…ホストインタフェース、103…マイクロプロセッサ、104…キャッシュメモリ、105…ドライブインタフェース、106…ファイバチャネルケーブル（第1のインタフェース）、106a…ファイバチャネルケーブル（インタフェース）、107…ディスクドライブ（第1の単体ドライブ）、108…外部ポート、109…ファイバチャネルケーブル（第2のインタフェース）、109a…ファイバチャネルケーブル、110…パネル、200…外部ディスク装置、201…外部ポ

ート、202…外部ポート、203…ファイバチャネルケーブル、204…ディスクドライブ（第2の単体ドライブ）、301…ドライブ管理テーブル、301-1-a…D_ID、301-1-b…ループID、301-1-c…WWN、301-1-d…副ドライブへのポインタ、301-1-e…ドライブ正常フラグ、301-2-a…D_ID、301-2-b…ループID、301-2-c…WWN、301-2-d…主ドライブへのポインタ、301-2-e…ドライブ正常フラグ、302…データ管理テーブル、302-a…キャッシュメモリ上のデータへのポインタ、302-b…D_ID、302-c…LBA、302-d…主書込未反映フラグ、302-e…副書込未反映フラグ。

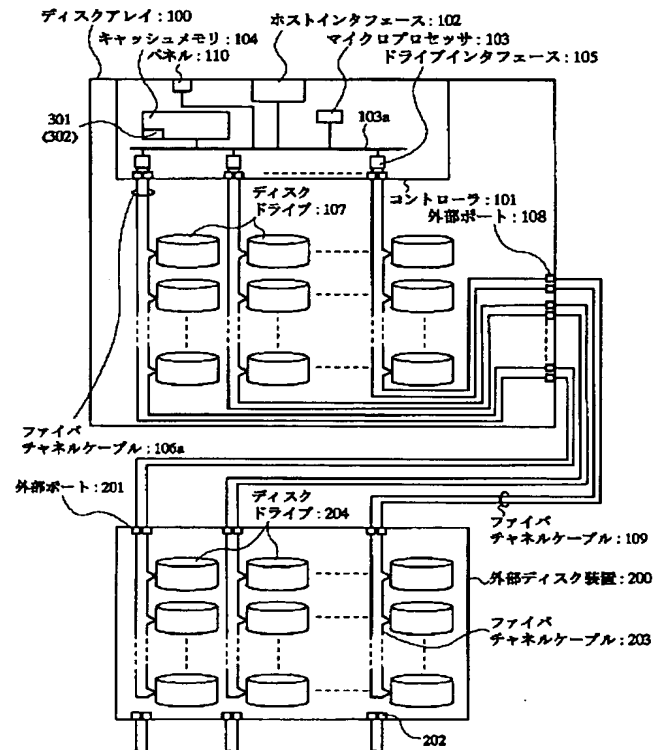
【図1】

図 1

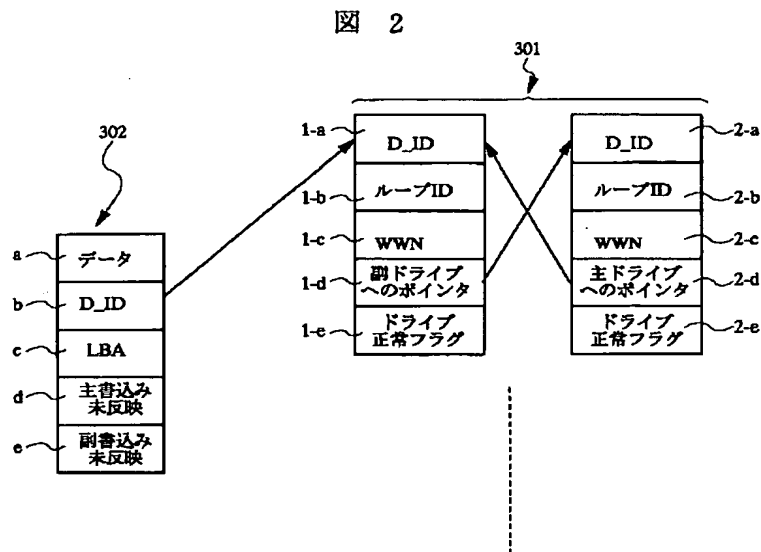


【図5】

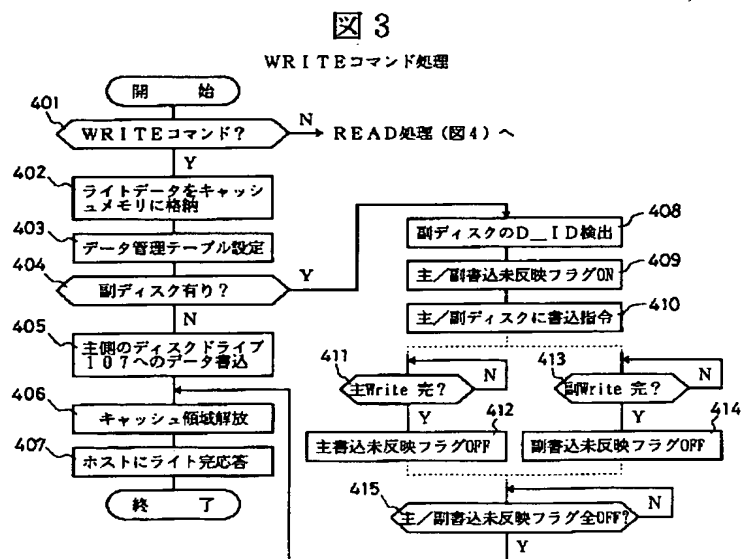
図 5



【図 2】



【図 3】



【図 4】

図 4

READコマンド処理

